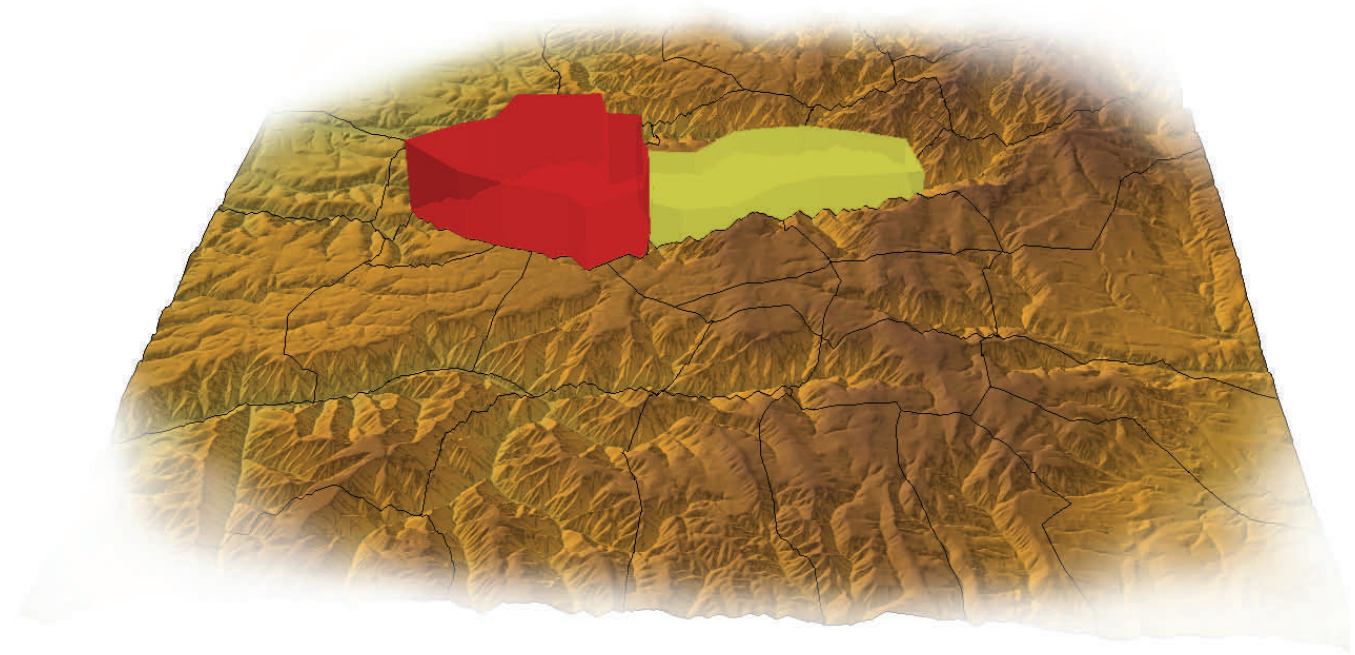


Approaches to Landscape-Scale Modeling



A hypothetical landscape with agency management units (e.g., county, wildlife management unit), over which data is frequently modeled or summarized at the landscape level. The red and yellow units could represent mean prevalence, animal density, or some other feature of the landscape.

The primary goals of landscape-scale modeling are identifying areas of high-risk for spread, as well as general patterns of disease spread. Another modeling goal is to link the spatial structure of wildlife populations and the spatial variability in abiotic and biotic attributes of their environment with disease transmission dynamics. Large-scale modeling often involves averaging over large areas to depict and analyze the patterns and spread of a disease. An inherent shortcoming of averaging is the loss of understanding of some biological processes, such as animal movements, on disease epidemiology. The term landscape epidemiology illustrates the concept by mapping a landscape in terms of spatial risk factors for infection and disease prevalence (Hess et al. 2002).

Features of the landscape and host that may affect disease distribution and transmission can be georeferenced and mapped. In addition to spatial variation, many factors vary in a temporal fashion, such as seasonally or annually. Both spatial and temporal data at the landscape scale can be useful in making predictions based on past conditions, and can be updated as conditions change or new information becomes available.

Statistical approaches seek correlations between environmental conditions and the distribution of disease, while mechanistic approaches attempt to identify biological processes that drive the observed patterns (Lawson 2001). The observed patchiness of a wildlife disease on a landscape could be the product of environmental factors that enhance the exis-

tence or transmission of the disease, or the distribution and movement patterns of hosts or vectors of the disease. At a landscape level, however, other factors that influence this dynamic relationship may emerge, such as effects of management intervention or localized human induced risk areas. Models at a landscape scale may allow us to tease out how these factors interact with the disease to produce the observed patterns.

CLUSTER ANALYSIS

Cluster analysis is a valuable approach for identifying disease patterns at multiple scales, but is especially useful for describing general point patterns across a landscape (Wakefield et al. 2000). From a decision-making standpoint, clustering at the landscape level can help to identify areas where disease incidence is higher than might

be expected by chance alone. In the regional-level section, we presented cluster analysis for multi-state or providence scales where data are likely to have coarse resolution, or visualization at large extents (i.e., multiple states or providences) requires a reduction in data accuracy and detail. At a regional scale, cluster analysis uses data that are either missing location information for a proportion of the data, or data are summarized into polygon count or prevalence data. Thus, prevalence, or counts of positive and negative cases, could be summarized over areas such as wildlife management units. Here we consider cluster analysis for data with higher spatial resolution, such as data sets where specific location information (e.g., UTM coordinates) are available for most of the samples. For these data, the assignment of a case to a cluster is based primarily on its coordinates. Specifically, CWD positive cases are placed in clusters based on their relative distance to other positive cases while account-

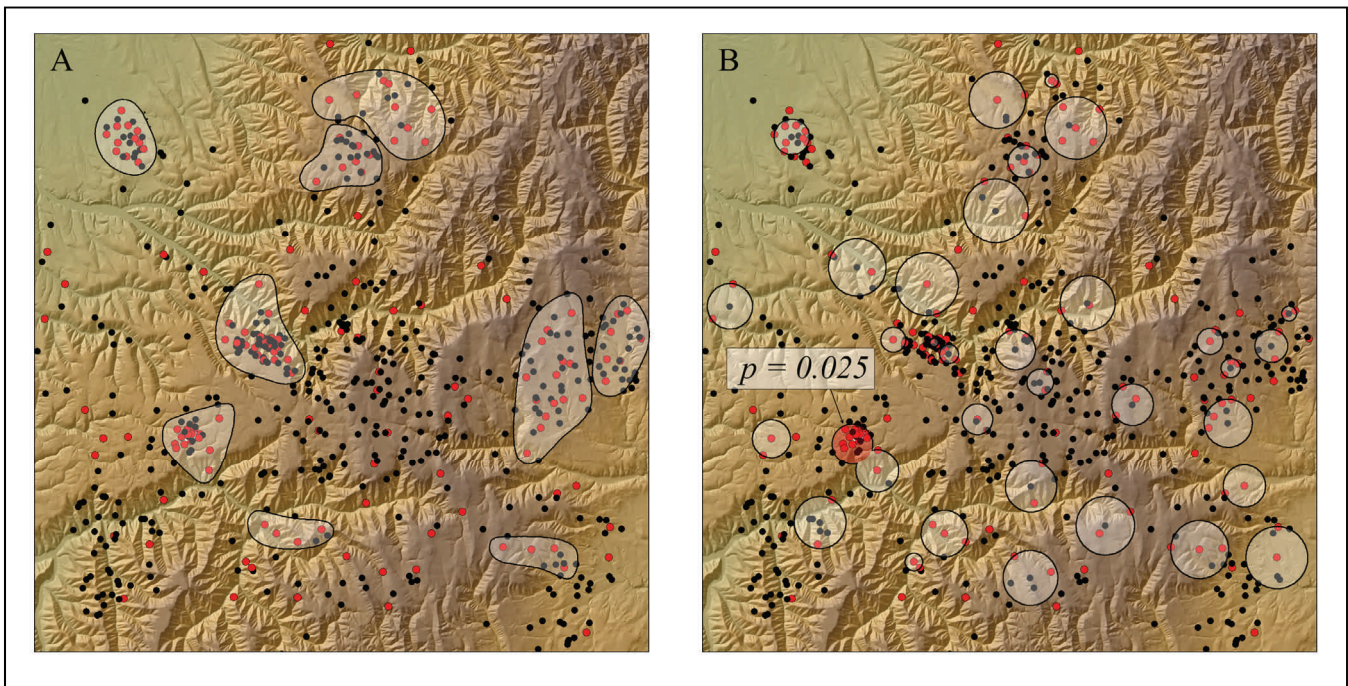


Figure 2.1. Simulated data illustrating the concept of residual spatial variation. Red circles represent positive cases and black circles represent negative cases. (A) A density based clustering algorithm that does not take into account the negative cases, but is well suited to detecting irregular shapes. (B) A spatial scan clustering algorithm which accounts for the underlying structure of the population but uses a circular (or elliptical) window to find clusters. The scan method (B) identifies many possible clusters (white-shaded circles), but only identifies one as statistically significant (red-shaded circle, $p = 0.025$). The cluster in the upper left was selected as the second most likely cluster, but was not significant.

ing for distances between negative cases. Although we discuss cluster analysis at the landscape scale, it could also be used at the regional scale if most samples have spatial coordinates.

In the context of spatial epidemiology, the definition of a cluster takes on rather specific meaning:

“A spatial aggregation of cases relative to the underlying distribution [locations] of non-cases, with a unique structure which differentiates it from other clusters.”

or

“the residual spatial variation in locations of disease (Diggle 2000).”

The important point is that both definitions explicitly take into account the spatial heterogeneity of cases relative to non-cases. This idea is particularly relevant to the study of CWD, where hosts or infectious agent may be patchily distributed in space, and data comes from a variety of sources (e.g., agency culling, hunting, captive cervid operations, etc) that also can be patchily distributed. Simple scatterplots of locations of CWD samples clearly show heterogeneity in the density of samples at multiple scales (Farnsworth et al. 2006). Thus, a simple clustering of positive cases (ignoring non-cases) can be misleading as apparent disease clusters may be explained simply by a clustering in sample collection density and represent sampling effort rather than disease clustering (Figure 2.1) (Kulldorff and Nagarwalla 1995). Accordingly, the inference of disease “hot-spots” via cluster analysis should only be considered when they represent, in Diggle’s (2000) words, “the residual spatial variation” of positive and negative samples.

One of the most widely used cluster approaches for epidemiological data that accounts for the underlying density of non-cases is the spatial scan statistic. Developed by Kulldorff (1997), it identifies a significant excess

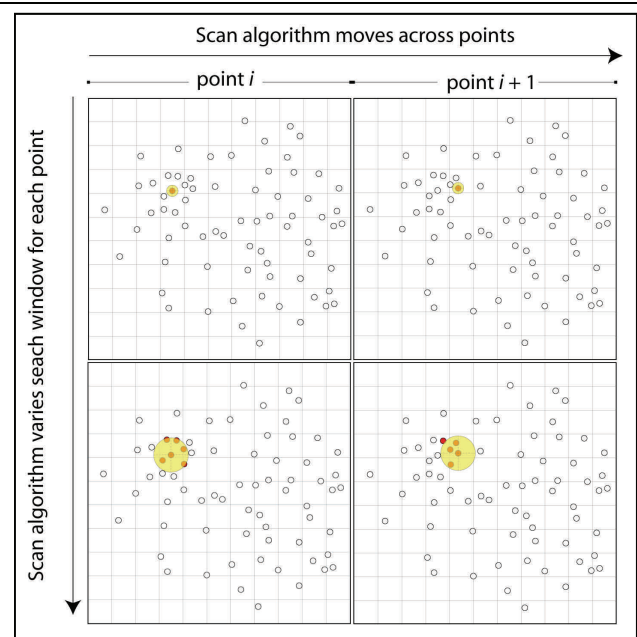


Figure 2.2. Shows an example of how the scan statistic method works. At each data point the algorithm samples the number of members inside a set of circle ranging from 0 to a user specified maximum limit (either an actual radius for the circle or a maximum proportion of the population). The algorithm then moves on to sample the next point in the dataset and continues this process until all points have been sampled.

of cases within a moving window that visits all spatial locations, increasing in size at each location until it reaches a predetermined upper size limit (Figure 2.2). The scan statistic provides a measure of how unlikely it would be to encounter the observed excess of cases in a larger comparison region. A more detailed description of the scan statistic is covered in the focal method of this section. Several variations of the scan statistic method have been applied to a wide variety of epidemiological studies for cluster detection including non-Hodgkin’s lymphoma, child mortality, and bovine TB (Perez et al. 2002, Sankoh et al. 2001, Tango and Takahashi 2005, Veil et al. 2000).

Clark and Larson (2002) group clusters into three classes: spatial clusters, temporal clusters, and spatio-temporal clusters. Purely spatial clusters represent risk that remains constant over time. Conversely, temporal clusters exist

across entire regions, but only for relatively brief time periods. Spatio-temporal clusters imply a temporary localized region of elevated disease within a larger space-time context. For example, spatio-temporal clustering methods were used to study outbreaks of acute respiratory disease in cattle herds in Norway (Norstrom et al. 2000).

For highly transmissible and fast moving infectious agents, the incorporation of time is often wise. In the case of the CWD, this is unlikely to be useful due its relatively slow rate of spread. The practical aspects of sample collection also influence the ability to incorporate time into a cluster analysis. Sampling often occurs in periodic “bursts” of relative short periods, such as focal (“hot spot”) culling over days, or as annual sustained events such as hunting seasons, which occur over days or weeks. Accordingly, we suggest that addition of time to a cluster analysis for CWD be restricted to annual periods. If little spread is suspected, data from multiple years can be combined for cluster analysis. To do this, data within years (and/or between years) are combined onto a map and considered only as spatial clusters. Attempts to assess changes in prevalence across years should be done carefully, as spatial patterns of culling and hunting may not be consistent across years.

Questions addressed / model predictions:

1. Identifies localized disease high prevalence areas or “hot spots”.
2. Identifies spatial disease patterns at different scales.

Data required:

1. Spatial coordinates of positive and negative samples if disease cases do not ‘present’ themselves – such as for CWD cases.
2. Random and representative samples.
3. Covariates (if applicable).

Output:

1. Assigns positive cases to a particular cluster.

ter.

2. Identifies potential spatial covariates to disease pattern.
3. Identifies high-risk areas and cluster maps.

General usefulness:

The main strength of cluster analysis is that it can be used to depict and describe spatial patterns of a disease and identify hot spots of high disease prevalence (at which management intervention could be targeted), as well as to estimate a risk surface. Cluster analyses are exploratory, but can be useful for hypothesis generation. For example, if a risk surface is generated, models with environmental/ecological (abiotic and biotic) covariates, genetic covariates, or other factors suspected to be related to the disease, can be evaluated. Cluster analysis is a valuable initial step in examining the spatial epidemiology of a disease.

Usefulness to CWD modeling and/or management:

Cluster analysis is a valuable descriptive tool for CWD surveillance data. At the landscape scale, we recommend that cluster analysis be conducted as a first step for georeferenced CWD surveillance data. The usefulness of cluster methods for CWD is the same as described above in “General usefulness”. We considered the special case of cluster analysis, the use of kernel density estimators, to generate a risk surface, in the regional-level section on Risk Analysis/Assessment.

GEOSTATISTICAL ANALYSIS



Geostatistics are statistics pertaining to the earth, or statistical techniques that emphasize locations with an areal (spatial) distribution. Geostatistics is usually concerned with statistical theory and applications for spatial processes which have a continuous (or nearly continuous) spatial index. In traditional statistical analysis we assume that observations are taken under identical conditions, and independently from one observation to another. However, with spatial

data, the location of each observation gives rise to spatial dependence and heterogeneity (Cressie 1993). Spatial autocorrelation measures the degree of this statistical association between data units for different distances, and if autocorrelation occurs it becomes important to describe the underlying spatial process (Diggle et al. 2003). For diseases, like CWD, positive spatial autocorrelation usually indicates areas of either high or low risk. Negative autocorrelation seems unlikely as it implies areas of higher disease are adjacent to areas of low disease. Such negative correlation patterns might imply landscape boundaries to disease spread.

Geostatistical methods provide an important approach for modeling and correcting the spatial autocorrelation that commonly occurs in disease data. The overall goal of a geostatistical analysis is to assess factors that may be associated with the occurrence of disease, while accounting for spatial dependency. For example, explanatory variables can be incorporated in the analysis to account for factors related to disease risk (males vs. females), disease exposure (age), or type of disease transmission (density-dependent vs. frequency-dependent transmission), as well as to evaluate other factors that might affect disease patterns (animal movement and habitat patterns). These spatial regression models take into account both the importance of dependent variables and spatial dependence. An advantage of the geostatistical approach is that it recognizes both larger scale spatial trends and local spatial correlations. Most CWD affected areas will have substantial small-scale variation, typically exhibiting strong positive correlation between data from nearby spatial locations.

The use of geostatistical methods requires observations of a response variable (0, 1 for CWD infection status) for individual animals or summary of prevalence in a small area, and the spatial locations of these responses. Ideally, locations should be truly continuous in space, but small scale clustering (summary) of data would not likely cause major violations of

this requirement. A geostatistical regression model looks generally as follows: $Z = X\beta + \delta$ (d), with fixed effect dependent variables X modified by β , and $\delta(d)$ is a zero-mean error vector that is spatially correlated (a function of distance) according to the model selected through a geostatistical analysis. In the case of disease, the response variable (Z) is typically discrete data (binary for disease status or Poisson for counts of infected animals). In particular, linear models may not always perform well and discrete models may be more appropriate for disease data. Some useful process models (commonly called link functions) include the binary (or logistic) model for CWD status (susceptible or infected), the Poisson for CWD prevalence data using counts of infected and susceptible animals, complementary log-log models for estimation of disease prevalence using age-prevalence data, and more complex hazard rate models (Heisey et al. 2006).

In geostatistical analysis, spatial variance is modeled using a parametric pattern (Cressie 1993) that best describes the spatial correlation (dependency) in the data. This variance can be considered to have both distance and directional properties, and it is evaluated using the variogram (or semi-variogram = variogram/2). The variogram is the cornerstone of geostatistical analysis, and is treated as a random process (variable). The variogram is used to assess the degree of variance between spatial locations as a function of the distance between locations (Figure 2.3).

The basic structural components of the variogram (Figure 2.4) include the parametric model's underlying spatial dependency and the estimated model parameters (nugget, sill, and range). The nugget measures discontinuity at the variogram origin ($h=0$ or minimum variogram lag distance). Theoretically, we expect the nugget to have a value of 0 because points are perfectly correlated with themselves. Actual data often show differences between data collected at very close locations in space. This variation usually results from random variation at scales below the minimum lag dis-

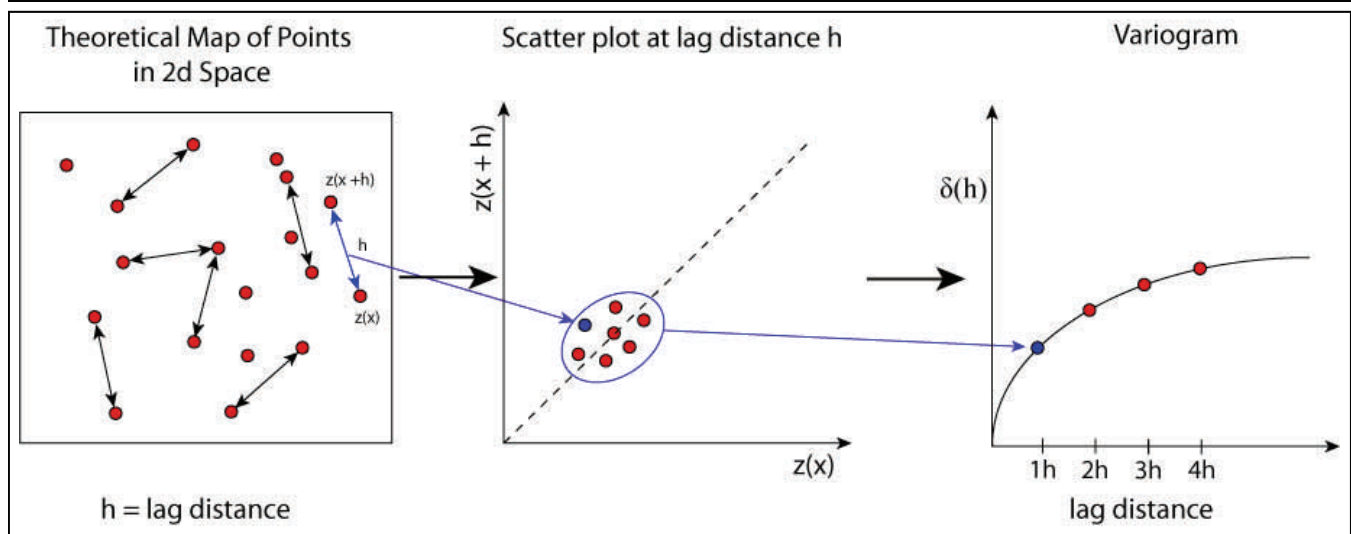


Figure 2.3. Calculation of the variogram. All pairs of points separated by lag distance h are identified (left). For each pair of points the difference between the values of the points is calculated (middle). For each lag distance the squared differences are summed and divided by twice the number of points for that lag distance (right). The variogram is created by repeating this process at different lag distances (e.g. $1h$, $2h$, $3h$, etc.). Adapted from Shibli 1997.

tance (microscale variation) or from measurement error. Thus the nugget is an estimate of sampling plus microscale variation. Variogram patterns (i.e., variance) generally increase with lag distances because close points tend to be more similar than those located farther apart. The lag distance at which the variogram values stabilize and the distance beyond which data appear to be independent is called the range. The final structural component, the sill, repre-

sents the value of the variogram at the range and indicates the total amount of correlated spatial variation in the dataset.

When the variogram exhibits different sill or range values in different geometric directions the spatial correlation is a function of both distance and direction, and the random process is considered to be anisotropic. This is usually caused by underlying processes evolving differently in space or underlying landscape features (i.e., valleys, ridges, soil). If the variogram is only a function of distance (i.e. the sills will be the same in all directions) it is considered to be isotropic. In the isotropic case, the pattern (linear, exponential, quadratic, wave, power, etc) of the variogram is used to identify and model the underlying spatial dependency in the data. This model of spatial autocorrelation, $\delta(d)$, is incorporated with the larger process model that accounts for other trends or covariate predictors of disease.

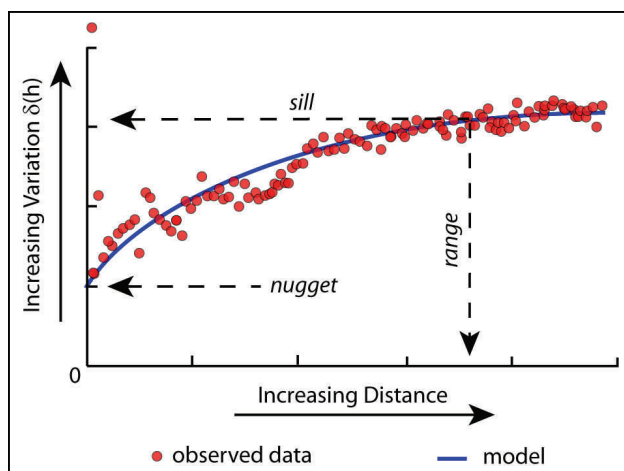


Figure 2.4. Variogram for an exponential model of spatial variation showing the nugget (random variation), sill (total variation), and range (distance of correlation). Adapted from USDA. F.S.—N.E.F.I.A.

Joly et al. (2006) conducted a geostatistical analysis of CWD infection data for south-central Wisconsin to evaluate both spatial autocorrelation and ecological factors hypothesized to be related to apparent prevalence. They used data to evaluate the potential effects of

deer habitat, age, sex, and distance and direction from a suspected introduction site as factors affecting CWD prevalence, which was aggregated at the section (i.e., 2.6 km² or 1 mi²) scale. Their analysis indicated that CWD prevalence declined over both a broad scale distance from the center of the outbreak area and at a scale reflecting local spatial correlation (i.e., 3.2 km or 2 mi radius). In addition, deer habitat was a significant predictor of CWD prevalence. Joly et al. (2006) used the resulting regression model to produce a map of predicted CWD prevalence.

Additional complexities may also occur in spatial disease data, but these topics are beyond the scope of our review. For example, CWD data may be aggregated into a finite collection of regular (e.g., sections, townships) or irregular (e.g., counties, wildlife management units) spatial sites or cells called lattices. Figure 1.3 illustrates the difference between continuous location of cases (dots) and potential aggregation into an irregular lattice of polygons. Methods for analysis of lattice data is described by Cressie (1993) and traditional applications to human diseases are considered by Elliott et al. (2001) and by Lawson and Williams (2001). It is also possible to consider spatiotemporal analysis, but in most cases this is simplified to a purely spatial process by aggregating over time. Because CWD is typically a slowly transmitted and slowly spreading disease it seems appropriate to aggregate over relatively short time frames (e.g., < 5-10 years). The general goal of geostatistical analysis is to develop models that incorporate disease risk factors and predict CWD infection at known spatial locations.

Questions addressed / model predictions:

1. Evaluates the spatial extent (distance) and direction of autocorrelation found in disease patterns.
2. Potentially evaluates the relationship of biotic and abiotic factors on disease infection or prevalence.

Data required:

1. Spatial coordinates of positive and negative CWD cases.
2. Spatial coordinates of polygon centroids if CWD cases are aggregated. However, for geostatistical analysis, aggregation should occur at a relative small scale compared to the area considered in the analysis.
3. Individual animal covariate data (e.g., age, sex) for factors of interest in predicting disease risk.
4. Spatial environmental or ecological covariate data (habitat, animal density, risk variables) for factors of interest in predicting disease risk.

Output:

1. Estimates the spatial autocorrelation of disease related to distance and/or direction.
2. Estimates prevalence parameters, covariate effect sizes, and other related statistics for factors/variables affecting the probability of disease.
3. Potentially depicts CWD spatial prevalence based on spatial autocorrelation and other significant factors affecting prevalence.

General usefulness:

Geostatistical methods are highly useful for spatial analysis of ecological and geographic processes that are sampled at irregular or random locations. These methods are most useful when the goal of the analysis is prediction at an unobserved spatial location. Geostatistical methods may be additionally useful when there is spatial dependence in the process that generates spatial patterns. Additional explanatory variables can also be included in the geostatistical analysis, leading to an investigation of spatial effects while controlling for explanatory factors or vice versa. Geostatistical methods generally assume a relatively small error in the spatial scale of locating animals (data points). This assumption may be reasonable when animals have small home ranges compared to the area of general analysis; however, for animals with seasonal migrations it may be important to separate analyses based on distinct summer or winter distributions. The geo-relational database structure of a geographic

information system (GIS) is ideally suited for storing and manipulating data used in geostatistical analyses.

Usefulness to CWD modeling and/or management:

If relevant spatial data are available, a geostatistical analysis has high potential applicability to CWD modeling and management. Based on the slow rate of CWD transmission and spread, potentially irregular spread of disease through heterogeneous habitats, and typically low rates of infection, it seems likely that spatial patterns and dependencies will be important components of CWD spatial distribution. Analysis of CWD data using geostatistical methods can facilitate the evaluation of biotic and abiotic factors affecting the risk of infection while concurrently accounting for likely spatial dependence. Modeling results can be used to produce maps of predicted CWD prevalence or risk. In addition, geostatistical analysis may provide useful insights on the dynamics of CWD spread by describing the extent of spatial correlation on the landscape. However, geostatistical methods may not be useful for all CWD infected areas. In particular, the spatial and temporal scales associated with data collection and CWD case location should receive careful consideration. Finally, aggregation of CWD data over a number of years seems highly likely to improve the distribution and precision of spatial data. Because CWD is a slowly transmitted disease, aggregation over a few years may not be problematic; however, there are currently no specific guidelines for determining the appropriate time frame for aggregation.

discrete (i.e., an integer value) and finite; thus in the color example there would be no continuous shading and each cell would be one of a finite number of possible colors. For a three-dimensional problem, the squares would be cubes and the analogy would be a Rubik's cube. For CWD, time could be the third dimension so that x and y represent spatial extent, and z represents temporal pattern. Now we come to the 'automaton' part. As a cellular automaton model runs through time, at each time interval the state of the cells can change, or not, based on a deterministic or probabilistic rule. To enact the rule, each cell looks at the states, or color in this example, of nearby cells, and its own state (color), and then applies the rule to decide its state (color) in the next time step. All the cells change at the same time. This collection of cell-states and rule-based changes is called a *cellular automaton*, or cellular automata model. Two-dimensional systems of grid-cells are also called lattice-systems. Cells need not be blocks but can be any arbitrary shape. Although time must be discrete, it can be at any interval, from sub-seconds to years or longer. Even with very simple rules for each cell, these models can result in complex patterns and dynamics.

Cellular automaton models have been used to study the spatial and temporal rates of disease spread in spatially distributed host populations, as well as to evaluate the effectiveness of vaccination intervention strategies (Rhodes and Anderson 1997). A probabilistic automata network SIS (susceptible-infective-susceptible) model was developed to evaluate the spread of an infectious disease in a population of moving individuals (Boccaro and Cheong 1993). When there was high movement, the spatial correlations in infection and recovery disappear and, as expected, the behavior of the system was then correctly predicted by a mean-field model, which assumed that every individual in the population is equally likely to contact every other individual. Results from the mean-field model diverged from the spatial cellular automata models when the neighborhood of interacting grid cells was reduced. At

CELLULAR AUTOMATA MODELS



To understand what a cellular automata is, consider a chessboard or checker-board. The cellular part is represented by the squares and each cell can have one distinct state, such as color. Thus, a cell could be red or black (2 states), or yellow, white, or orange (3 states). The state must be

a fine scale, cellular automata models can be individual-based models where cells can represent individuals with a spatial state. At a larger-scale, cells could represent populations or spatial areas, which would be more appropriate for landscape-level modeling.

Because cellular automaton models are general, they can be used to address a huge range of questions. The level of detail, spatial and temporal resolution, and inputs and outputs can be adapted to the specific questions of interest. Because this class of models includes such a broad range applications, we do not address these categories below.

General usefulness:

Cellular automaton models are particularly useful for developing an understanding how different factors may affect the spatial spread and distribution of a disease at a variety of spatial scales. The main strength of this method is its flexibility and the potential simplicity of rules within a cell. However, realistic data on vital rates, movement, transmission dynamics, population structure and distribution, relevant environmental/ecological (biotic and abiotic) factors, etc. are required to parameterize realistic cellular automaton models. Cellular automaton models are as realistic as the data used in their construction. Still, even with incomplete data, they may offer heuristic insights into the disease system.

Usefulness to CWD modeling and/or management:

The cellular automata method is very general, and could be used for a variety of different purposes. Given the appropriate data, these models could be used to investigate the likelihood of different localities as the point of CWD origin. The usefulness of this approach will depend on how a model is constructed and whether there is adequate data to parameterize it. However, the cellular automata umbrella offers a viable and potentially effective approach to spatial modeling of CWD to researchers with clear questions and appropriate data. Because of its potential usefulness for

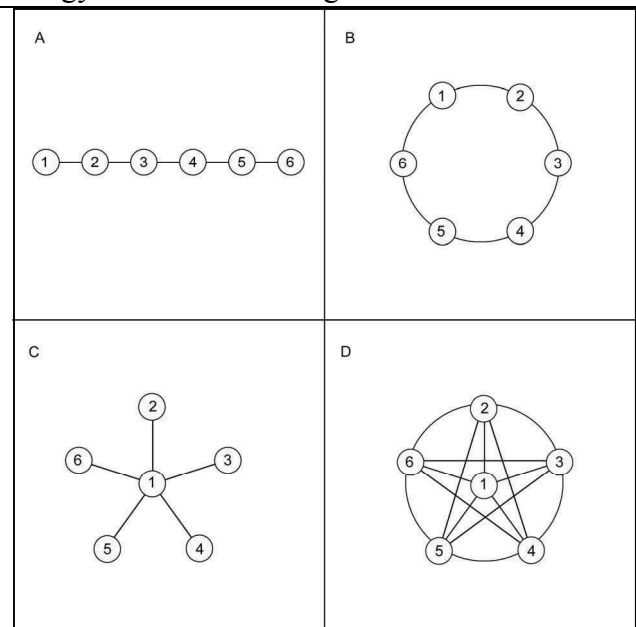



Figure 2.5. Different spatial configurations and continuity used in metapopulation modeling: (A) chain or necklace model, (B) loop model, (C) spider model, and (D) island model. Lines show connectivity between patches via dispersal.

modeling CWD spatial epidemiology, we present a related individual-based model as the focal approach for the fine-scale section.

METAPOPULATION MODELS

 Metapopulation models are based on the premise that there is a set of populations distributed over a number of patches, or areas, which are connected by dispersal (Figure 2.5). Early work on the theory of metapopulations assumed that dispersal among patches was limited and patches would go extinct and/or become recolonized over time (Levins 1969, Hanski and Gilpin 1991). A premise of metapopulation models is that patches “wink on and off” between an occupied and unoccupied state. With diseases, this would mean that the disease would wink on and off in the populations. However, this approach could be adapted for diseases that arrive and either do not wink off (population unoccupied by disease) or wink off very slowly. Thus, we prefer to think of meta-

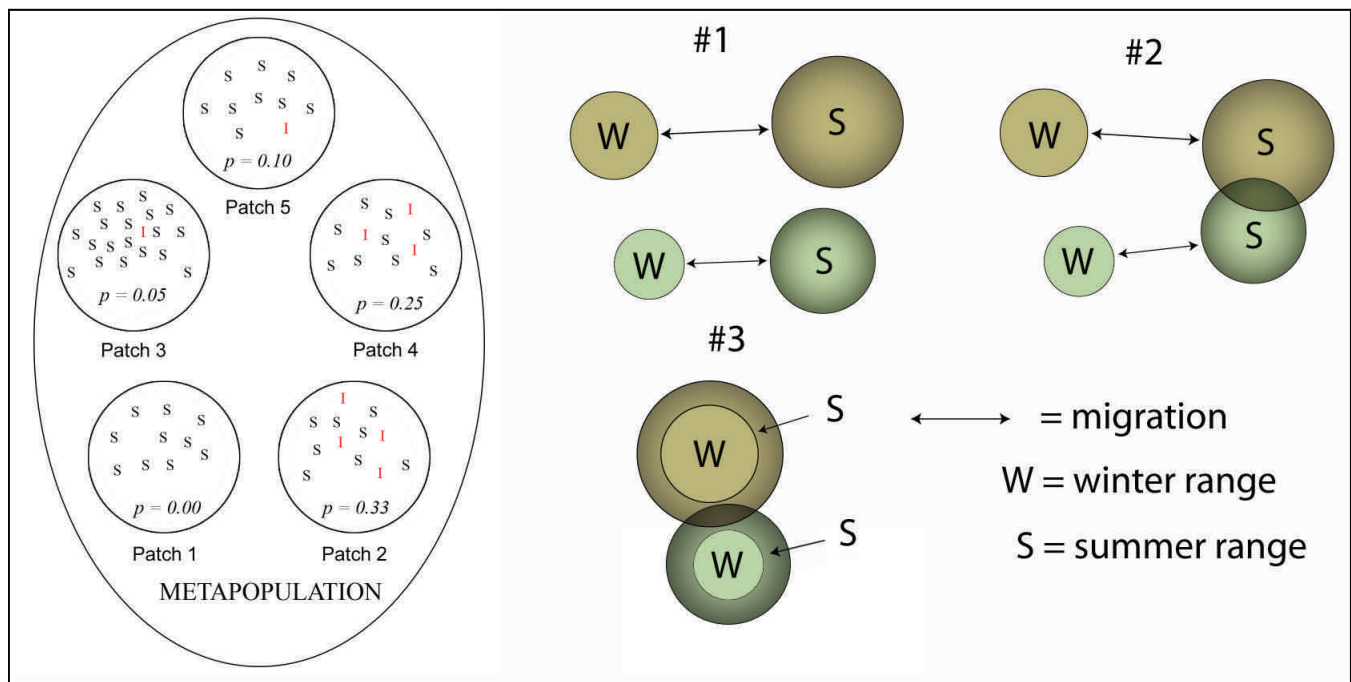


Figure 2.6. Depiction of infected (I) and susceptible (S) individuals within metapopulations, adapted from Hess (1996), and the adaptation of this model to CWD for mule deer. Traditional metapopulations are connected by dispersal. For mule deer, the mechanism of connection between populations may also be due to either migration to summer range combined with summer range overlap or summer range expansion and overlap. Case # 1 there is no overlap, and hence is connection or metapopulation dynamics. In case # 2 there is migration to summer range combined with summer range overlap. This situation, the product of the probability of migrating \times probability of range overlap could represent connection between the populations, which is analogous to the dispersal probability in metapopulation models. In case #3 there is no migration, but there is range expansion in the summer such that there is overlap between the 2 populations. In this situation, the probability of range overlap could represent connection between the populations. Finally, prevalence on winter range can be substituted for the numbers of I and S individuals from the metapopulation to adapt the metapopulation disease model of Hess (1996) to CWD models.

population models as part of a continuum of spatially (or socially) structured subpopulations. At one end of the continuum dispersal is very rare and subpopulations are relatively independent of one another. While at the other end of the continuum dispersal is frequent and subpopulations are more connected with one another.

Metapopulation models of single species are often used to investigate how dispersal and subdivision affects spatial and temporal population dynamics. However, these models can also be applied to wildlife disease systems (Hess 1996, Swinton 1998, Cross et al 2005) where patches are then defined as single hosts

or a group (herd) of hosts, which are colonized by the parasite (Hess 1996, Hess et al. 2002). Here the 'patch' refers to the population (group) of susceptible and infected animals inhabiting an area that is separate from other populations (groups), but connected by dispersal or movement. The probability a population is 'occupied' by a disease can be modeled following the traditional metapopulation approaches of site occupancy by a species.

Metapopulation models are relevant to modeling diseases when there is spatial (or social) structure in the host population such that host vital rates and disease transmission rates depend upon local conditions within the population (herd). This is often the case for wildlife

diseases that are directly transmitted because hosts are often structured into groups and individuals are most likely to be infected by others within the same group. These models commonly add complexity by explicitly modeling dispersal and connections among subpopulations to more realistically model heterogeneous mixing between populations, which improves prediction of disease spatial spread. Using these models researchers can ask questions about the spread of disease from one population to the next and the likely effectiveness different management strategies, such as quarantine that may be implemented in some subpopulations and not others (Figure 2.6).

Questions addressed / model predictions:

1. Estimates the probability a patch (e.g., population, subpopulation, individual, spatial area such as winter range or wildlife management unit) becomes infected or recovers from infection as a function of within and between population dynamics and movements.
2. Facilitates evaluation of management strategies that may be implemented spatially (e.g., ring vaccination or depopulating areas/populations with high prevalence).

Data required:

1. Dispersal or migration routes and probabilities of connection between patches.
2. For a detailed model, vital rates of population dynamics as well as transmission dynamics within each patch/population.

Output:

1. Estimates disease prevalence for each patch over time.
2. Estimates colonization and extinction probabilities of disease infection for each patch.
3. Estimates the probability that the disease goes extinct through time for the entire metapopulation.

General usefulness:

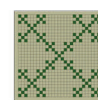
Metapopulation models could be useful for spatially structured populations to evaluate dif-

ferent management strategies, such as quarantine. This approach is also useful to explicitly model between-population or subpopulation spread of a disease, as opposed to within-population spread, which is modeled by many spatial epidemiology models.

Usefulness to CWD modeling and/or management:

One of the difficulties in applying metapopulation models to wildlife populations is the problem of defining a subpopulation. Subpopulations or herds may vary in size, location and the amount dispersal between herds over time. For example, grouping behavior of elk and deer in many areas of North America are likely to vary between summer and winter months (Conner and Miller 2004). In many cases the amount of movement among groups and the degree of independence among groups is unknown. Metapopulation models may still work where there is overlap between subpopulations as long as there is little mixing. The connection between subpopulations, which is modeled as dispersal probability in traditional metapopulation models, could be modeled by probability of exchange or other biological surrogate for connectivity. Metapopulation models are less useful for modeling the spatial epidemiology of CWD in the more contiguously distributed white-tailed deer populations that also lack seasonal movements between discrete summer and winter areas.

DIFFUSION MODELS



Diffusion models are based on an assumption that the process being modeled can be approximated by random motion. The rationale for using a diffusion model is that, although individuals do not move randomly, the collective behaviors of a large number of individuals cannot be distinguished from predictions of a diffusion approximation (called mean field approximation). This assumption vastly simplifies both the construction and evaluation of models. In the realm of disease ecology, these models pre-

dict the spread of disease over time and, as discussed here, also have a spatial dimension. Diffusion models have a rich history in ecology and they have been applied to an exceptionally wide range of processes, including spatial epidemiology of disease (Okubo 1980).

Because of the underlying assumptions, diffusion models are most often applied to geographically widespread diseases and when transportation of animals or disease agents by humans, or dispersal is unimportant. An early application of diffusion theory was Noble's (1974) model for the spread of bubonic plague in Europe. Strict diffusion models may provide insights to broad-scale processes, especially as an alternative comparison with more complex spatial models. Recently, Reluga et al. (2006) constructed models that combined mathematical advantages of a diffusion approximation while permitting the inclusion of spatial structure, including movement of an animal within a home range.

For many diffusion models, the first date an infected animal is reported for a given spatial area, such as a county or wildlife management unit, is the required data. From this data, a differential model or trend surface for rate of spread can be retrospectively fit. In this section, we describe 2 potentially useful forms of the diffusion model. The first, a semi-diffusion model, models disease spread from area to area rather than across continuous differential space and relaxes the assumption of random movement. The second is trend surface analysis, which is based on date of infection, and can retrospectively describe the spread of a disease in both space and time and identify likely corridors or barriers.

SEMI-DIFFUSION MODELS



We use the term 'semi-diffusion' because at a landscape scale transmission takes place across borders of areas, rather than from a point source outward. Semi-diffusion models may

allow more realistic modeling of spatial heterogeneity in disease spread because they summarize prevalence within area but not across areas, which requires data at the resolution of the area. In contrast, traditional spatial diffusion models attempt to model prevalence as continuous over space, which requires data at a fine resolution. Semi-diffusion models can be constructed for the spatial scale at which management is enacted. For example, wildlife diseases are often monitored by wildlife management units, and increased harvest or other management intervention actions are enacted for wildlife management units. With a semi-diffusion model, disease spread would be modeled across wildlife management units. The rate of diffusion or spread from one area to the other can be modeled as a function of prevalence in adjacent areas, proportion of border shared with infected areas, amount of connection via migration and dispersal, number of feeding sites in the area, etc. The diffusion rate could also be modeled as a function of any spatial environmental/ecological (biotic or abiotic) factors. A fundamental aspect of a semi-diffusion approach is explicit modeling of spatial dynamics (e.g., proportion of adjacent areas infected, prevalence in adjacent areas, etc.) in the difference equations, rather than modeling the effect through correlation structure. This approach could be framed as a model selection problem, with models representing different hypotheses about the spatial spread of the disease. Semi-diffusion modeling could be developed into a viable approach for any disease for which cases do not present themselves but where surveillance samples of infected and non-infected animals are collected.

Questions addressed / model predictions:

1. Estimates the probability an area becomes infected as a function of observed patterns, management actions, or environmental/ecological (biotic and abiotic) variables.
2. Predicts future spread of a disease from area to area.

Data required:

1. Date of first infection for a given spatial area, such as a wildlife management unit (date first detected is not the same as date of first infected, but if the two are close then date of first detected can be used for date of first infected).
2. Proportion of adjacent spatial areas that are infected for each time step.
3. Proportion of shared borders.
4. Prevalence within each area at each time step.
5. Proximity of spatial area to geographic features that could influence rate of spread, such as rivers (barriers or corridors), major highways (possible barriers), high ridge-lines (possible barriers), etc.
6. Covariates for a spatial area expected to influence rate of spread, such as human density, animal density, disease prevalence, or environmental/ecological (abiotic and biotic) variables.

Output:

1. Predicts the probability an area will become infected at a given time step.
2. Estimates spatial and temporal prevalence of each spatial area at each time step.
3. Predicts rate of spread across entire study area.
4. Can provide estimates of covariate effects.

General usefulness:


Semi-diffusion models could be quite useful for diseases that do not present themselves, and it could be useful for fast or slowly spreading diseases. It would also facilitate evaluation of the importance of environmental/ecological (abiotic and biotic) factors in the spread of a disease. The main disadvantage of this method is that it requires several years of adequate (i.e., enough samples to have a high probability of detecting the disease if it is present) surveillance samples from contiguous areas included in the model.

Usefulness to CWD modeling and/or management:

Data collected for CWD surveillance is typically of the type required for a semi-

diffusion model, making this approach potentially viable. Running this type of diffusion model backwards in time may help researchers and managers generate hypotheses about the factors important to CWD spread and potential originating areas. Thus, a semi-diffusion model may provide heuristic insight to understanding the present spatial patterns of CWD. However, a semi-diffusion approach has limited potential for predicting the probability of spread of CWD into uninfected areas. The problem is that to generate a good model from which to predict from, we would have to know how the disease spread. Because most new cases of CWD have not initiated from spread of the disease, but rather from increased surveillance of an area, there are very limited data on the temporal aspect of spread, and this temporal aspect is an essential element of this method or any diffusion method. Thus, we conclude this method is not useful for CWD modeling at the present, but may be in the future.

TREND SURFACE

 Trend surface analysis uses spatial polynomial models that accounts for global effects and local autocorrelation. Models are fit using least-squares regression, except that residual autocorrelation is included in the model. For this discussion based on Figure 2.6, we use months as the time interval and counties as the spatial areas of interest. A trend surface estimates contours for the number of months to the first reported case (Figure 2.6) using the centroid coordinates of wildlife management units or other relevant management area (Moore 1999). From these contour lines, partial differential equations ($\partial \text{time} / \partial x$, $\partial \text{time} / \partial y$) are derived to estimate the slope vectors from contour to contour. These slope vectors represent the rate of spread across the landscape. Large slopes represent high rates of spread which are interpreted as corridors. Similarly, areas with low rates of spread are interpreted as barriers (Figure 2.7).

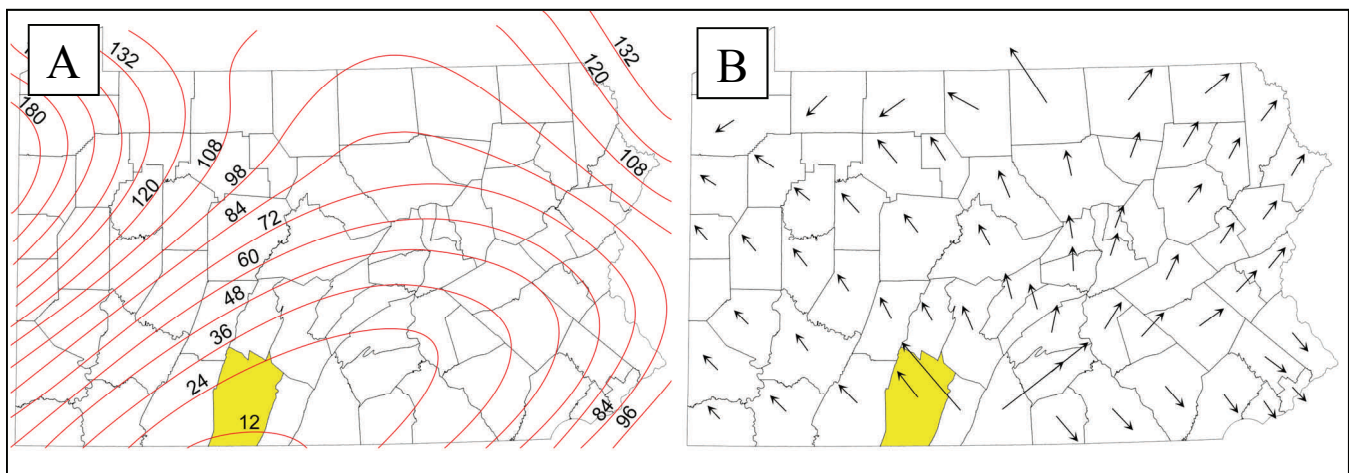


Figure 2.7. (A) The predicted trend surface for month to the first report of a raccoon rabies case of disease outbreak by county, Pennsylvania, USA, 1982-1996. This figure illustrates the general direction and movement of the diffusion process originating in the county shaded in yellow. The contours represent the predicted number of months to the first reported case. (B) Velocity vectors (derived from the partial differential equations from the trend surface) show the speed and direction of diffusion at each county centroid. Longer arrows represent faster spread. Adapted from Moore 1999.

Questions addressed / model predictions:

1. Estimates the rate of spread as a function of date that an area was first infected.
2. Identifies areas with fast and slow spread.

Data required:

1. Date of first infection for a given spatial area, such as a wildlife management unit (date first detected is not the same as date of first infected, but if the two are close then date of first detected can be used for date of first infected).
2. Centroid coordinates of the spatial areas.

Output:

1. Predicts contours of months to first reported case of disease.
2. Estimates rate of disease spread.

General usefulness:

Trend surface analysis is useful for infectious, moderate to quickly spreading epidemics in which cases present themselves. It is unlikely that the “first” case will be identified in an area, particularly in wildlife disease systems. Thus the rate of spread will be underestimated. However, as long as the time at which ‘first’ cases are identified is not biased by area (e.g., bias could arise if cases near ur-

ban areas are identified more quickly than cases in remote areas), trend surface analysis is usable for determining corridors of rapid spread and barrier areas which slow spread. Assigning areas to contain urban and remote areas could be important to avoid bias (e.g., incorrectly estimating spread to be fast around areas where there is a bias toward quick detection of the disease). However, this issue will be problematic when detection probabilities vary over space and time, which is likely for CWD.

Usefulness to CWD modeling and/or management:

Trend surface analysis is probably not useful for most CWD data sets because the date that an infected animal is first reported in an area may be completely unrelated to the date when the disease first occurred in the area. We note that the reason this method is not applicable is because of CWD surveillance and detection issues and not with the inherent usefulness of the method for CWD epidemiology. That is, if surveillance data were collected randomly and representatively over the entire state or area of interest, trend surface analysis could work well for CWD.

FOCAL APPROACH: CLUSTER ANALYSIS

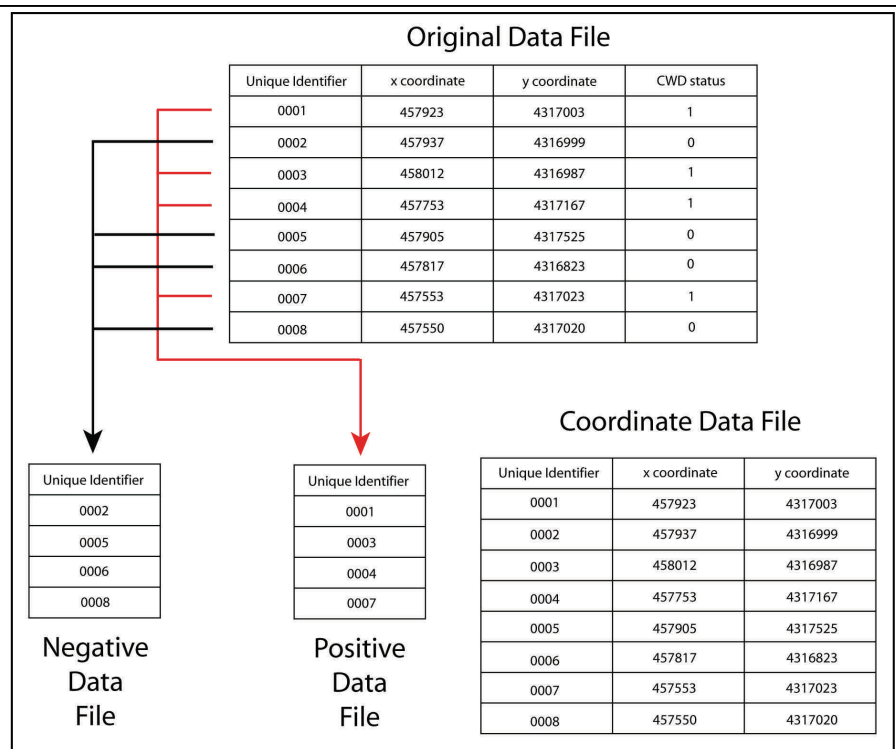
We selected cluster analysis as the focal approach for landscape level methods. While cluster analysis may not be as complicated or elegant as other approaches in this section, pattern detection with clustering remains a key element for both hypothesis formation and decision-making processes (Jain et al. 1999). State- or providence-wide surveillance programs, especially those involving hunter harvested animals (i.e., hunter harvest check stations) can provide large amounts of data (e.g., >25,000 samples) in a relatively short period of time (e.g., 1-4 months). Such a large influx of data often creates confusion and cluster analysis provides a convenient ‘entry-point’ into such a complex dataset. Further support for the use of clustering comes from situations where little prior information exists about the data (e.g., when starting a surveillance program, or monitoring new areas). In such cases it is prudent to begin with an exploratory analysis that makes as few assumptions about the data as possible - cluster analysis excels in this situation.

We selected Kulldorff’s (1997) spatial scan statistic for our clustering algorithm as it requires fewer *a priori* parameters (e.g., the number of total clusters or the number of individuals in a cluster) than most clustering methods. It has also been shown through power comparisons to be the most powerful method for detecting localized clusters (Tango and Takahashi 2005) and has been used to identify areas of high CWD prevalence from surveillance data on white-tailed deer (Joly et al. 2006).

Scan statistics work by moving a window of variable size across the points in a dataset, counting the number of observed and expected cases falling within the windows (see Figure 2.2). The most likely cluster of high prevalence (i.e., not occurring by chance) is determined using a maximum likelihood ratio statistic, which determines whether there is higher prevalence inside the window compared with outside. By maximizing the likelihood function over all locations and window sizes the most likely cluster is identified. Cluster specific p-values are obtained using Monte Carlo simulations for primary and secondary clusters. Scan statistics are appropriate for detecting clusters

Figure 2.8. Partitioning of a dataset for Bernoulli analysis in program SaTScan™. The original data array contains a unique identifier, spatial references (x/y) and disease status. This dataset is split into three individual files:

1. A vector of ID numbers for positive cases (red arrow).
2. A vector of ID numbers for negative cases (black arrow).
3. A coordinate array containing all records.



in space, time, or space-time, however we chose to demonstrate the method in the spatial dimension only. The following analysis was performed using the program *SaTScan*TM (Kulldorff 2006), which is available for free from <http://www.satscan.org/>.

Step #1: Data structure

Our example assumes that we are interested in detecting clusters of high prevalence of CWD by using disease status of individual animals, and therefore makes use of Bernoulli model framework. This type of model in *SaTScan*TM requires the spatial location of animals (x/y coordinates) and their status (0 for CWD negative, 1 for CWD positive). The *SaTScan*TM software requires partitioning of a dataset into three separate files; a positive case file, a negative case file, and a coordinate file covering all points. The positive and negative data files are simply a list of the respective unique identifiers. The coordinate file is linked to the two other files by the unique identifier field (Figure 2.8).

Step #2: Spatial scan & output data

We used two simulated data sets, each consisting of 1000 sampled individuals, to highlight the important aspects of the spatial scan statistic. The first dataset has a 10% prevalence rate randomly distributed across a theoretical landscape (Figure 2.9).

Three tabs are available (INPUT, ANALYSIS, OUTPUT) for the user to specify the model. For the Bernoulli approach we only need to specify the upper limit of the window size. This can either be set as a percentage of the population (both positive and negative events) or as the radius of the circle. For our analysis we used the default of 50% of the population. That is, the scan window will increase from 0 to a size that contains 50% of the population (i.e., 500 individuals) for each point it searches. With very large data sets searching 50% of the population for each point can take substantial time and we suggest setting the upper limit as a fixed radius rather than percentage of the population. The *SaTScan*TM output

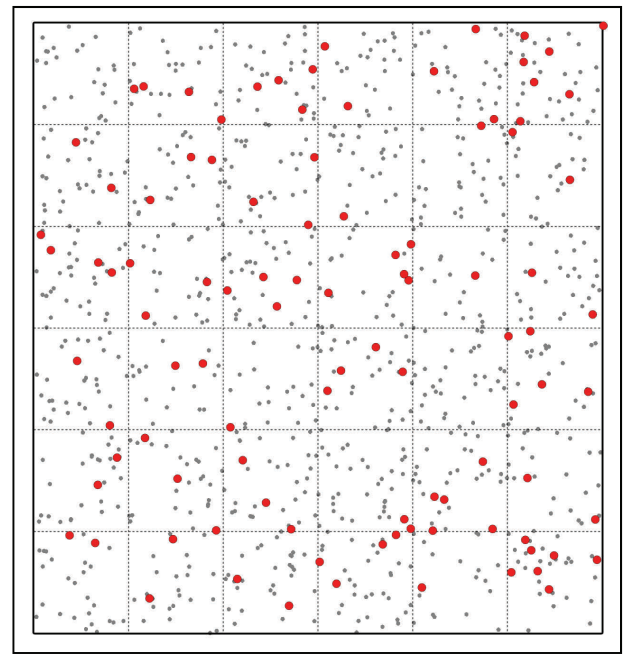


Figure 2.9. Randomly distributed 10% prevalence data. Red dots indicate positive cases ($n=100$), gray dots are negative cases ($n=900$).

provides various options to facilitate integration with a GIS mapping environment. This allows for visual mapping of the most likely clusters selected from a particular analysis.

It is important to remember that almost all clustering methods will produce clusters (that's their job!), even if they are not biologically relevant. To prevent such "false clustering" the spatial scan statistic tests if the pattern of positives and negatives inside any potential cluster is significantly different from the pattern observed outside that cluster. For our purely random 10% prevalence dataset, the Bernoulli spatial scan results show that even the most likely cluster is nonsignificant ($p = 0.636$) (Figure 2.10). We can infer there is no significant clustering of prevalence in this dataset.

To highlight the scan statistic's ability of detecting true clusters of higher prevalence from clusters caused by increased sampling intensity, we simulated a second dataset. This dataset contains an area of high prevalence and a second distinct area of dense monitoring (many positives coupled with many negatives)

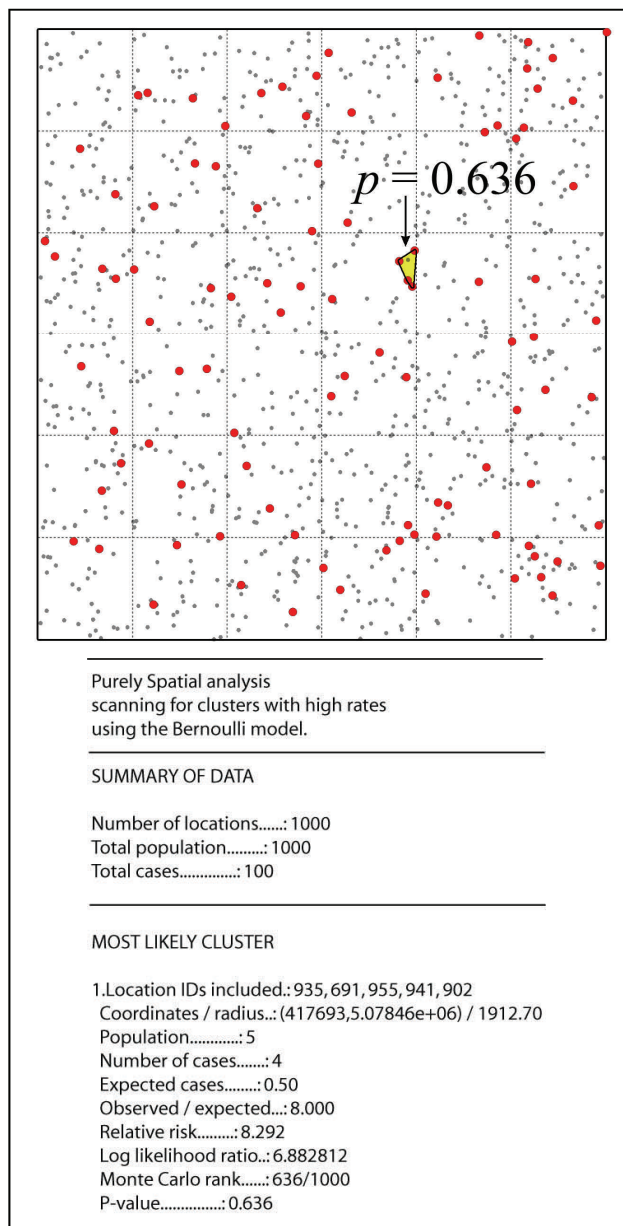


Figure 2.10. (Top) Cluster map for random 10% prevalence data. (Bottom) Output file from SaTScan™.

(Figure 2.11). Distinguishing between these two potential clusters represents the usual situation where simple spatial clustering often fails (by not accounting for the negative points). Simulating an area of dense monitoring is also relevant to CWD monitoring, because surveillance effort often varies across a landscape.

The scan statistic correctly identified the high prevalence area as significantly different

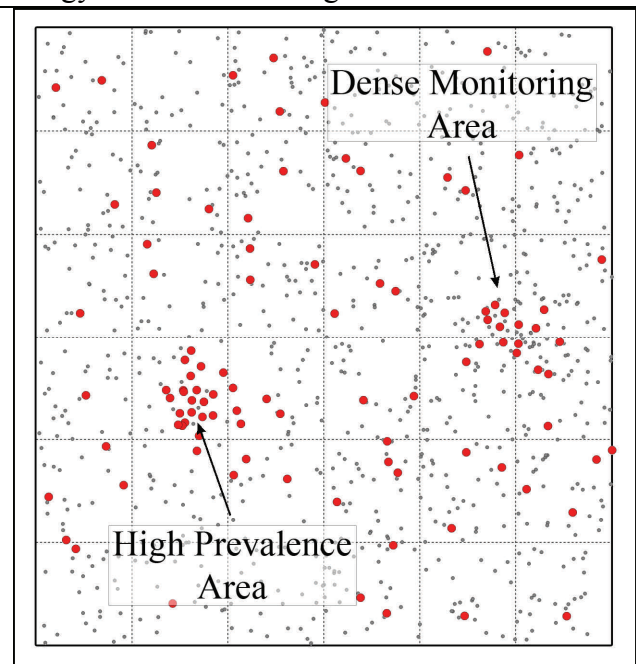


Figure 2.11. Clustered 10% prevalence data with high prevalence and dense monitoring regions. Red dots indicate positive cases ($n=100$), gray dots are negative cases ($n=900$).

from the surrounding population ($p = 0.001$). The increased sampling area was identified as a secondary cluster, but this cluster was not statistically significant. Accordingly, we can conclude that this cluster of positives is not significantly different from the surrounding population. It is simply the manifestation of a higher density of samples (both positive and negative) (Figure 2.12).

GENERAL CONCLUSIONS ABOUT LANDSCAPE-SCALE APPROACHES FOR CWD MODELING

As at the regional scale, we recommend cluster analysis as the first step of describing CWD patterns. Although there are many approaches to mechanistically modeling the spatial epidemiology of disease at this scale, their utility in the context of CWD is limited by several factors. The most important factor is lack of data. That is, we cannot reliably predict the spatial spread of CWD because sampling is usually insufficient to determine whether the disease is absent from some areas, or present but not detected. Diffusion models are inappro-

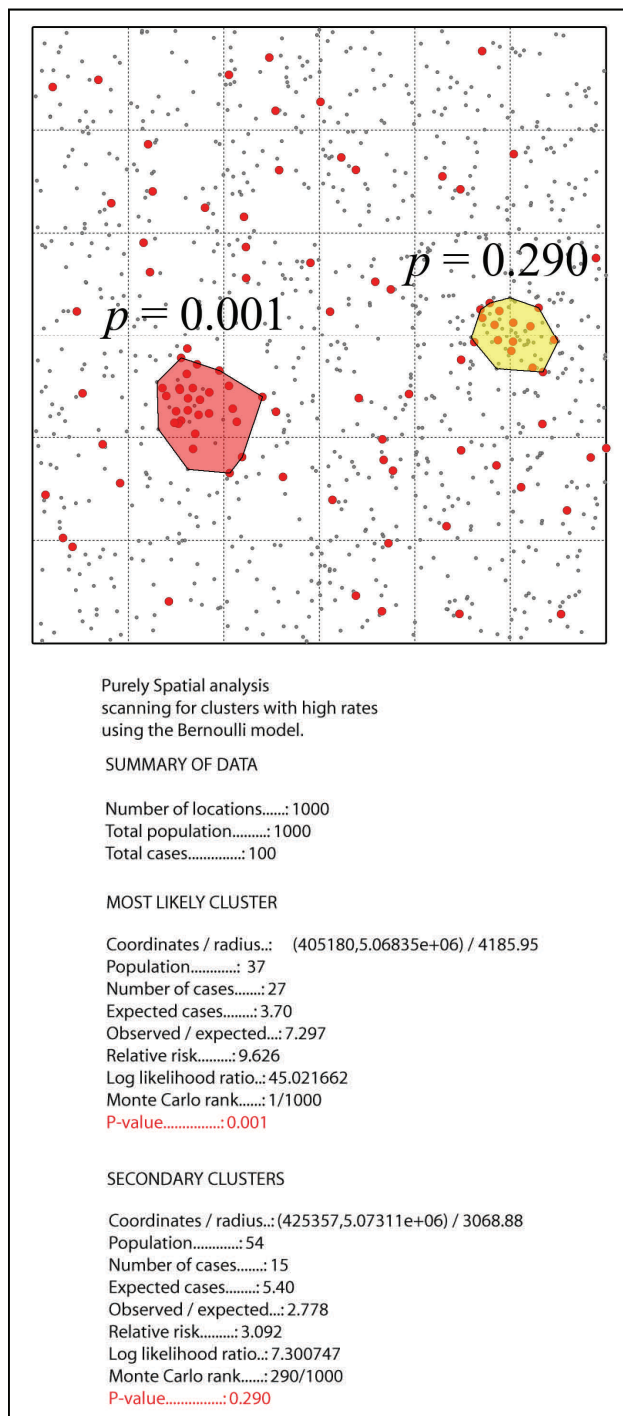


Figure 2.12. Spatial scan cluster map and SaTScan™ output table.

priate for additional reasons. For example, diffusion models are typically predicated on random mixing of individuals and, for some models, on the fact that the date associated with the first infected sample in an area represents the first occurrence of the disease in the area. The

random mixing assumption can be relaxed if factors causing differences in the rate of spread are explicitly included in the model. However, the second assumption is more problematic for CWD. The date when CWD is first detected in a new area usually does not represent a first infection, but rather a first detection. Diffusion methods typically assume that disease cases “present themselves” because they were developed for human diseases, where sick humans “present” themselves to doctors for treatment. However, deer with CWD rarely present themselves for testing. Thus, when we detect CWD for the first time in an area it likely that the disease was already there, perhaps for some time, but the area was not adequately sampled. Thus, until all areas are adequately sampled, surveillance data cannot be used in diffusion methods to model the spread of CWD in a meaningful way.

In addition, due to the apparently slow spatial spread of CWD, we believe that diffusion, cellular automata, and metapopulation models are unlikely to be particularly useful for addressing many CWD questions. In particular, we expect that the ecological system could change dramatically (e.g., hunting pressure, land use, water distributions, animal translocations [legal or otherwise], etc.) over the amount of time that it is likely to take CWD to spread, and that these factors might play a much larger role than the diffusion process. We have one caveat to our skepticism. If surveillance data over larger areas are adequately sampled, the semi-diffusion approach may be useful for predicting potential spread, or to evaluate management actions occurring at an appropriate scale, such as county or wildlife management unit. However, given that CWD appears to spread slowly, even if the semi-diffusion approach is viable it may not yield helpful results. That is, it may be years before CWD is predicted to move from populations in one large area to populations in another large area.

The largest data gap at the landscape scale is the lack of adequate sampling to detect

newly infected areas. Enough samples to ensure a 99% probability of detecting a prevalence of <1% should be collected over all areas of interest in order to predict any future spread of CWD. Sample sizes and designs to achieve this were thoroughly discussed in the previous workshop (Samuel et al. 2003). Because of inadequate samples, the first time CWD is detected in an area often represents the first time there are adequate samples and power to detect low prevalence CWD, not the first time it occurs or “spread” there. Modeling spread of disease based on observed patterns will not be valid until there are adequate samples in the relevant study areas.

In areas where sampling is powerful enough to describe present spatial patterns of CWD epidemiology, running models backwards in time may be a fruitful line of future investigation. This approach could reveal likely originating locations and times, as well as potential patterns of spread, that led to present patterns. Hypotheses of originating locations, times, and patterns of spread, including corridors and barriers, could be constructed. The forward projection of the outcomes of these hypotheses could be compared to the observed patterns via model fit statistics. The endemic area of Colorado and Wyoming may be an area with adequate samples to attempt a backwards time approach.

LITERATURE CITED:

- Boccara, N., and K. Cheong. 1993. Critical behavior of a probabilistic automata network SIS model for the spread of an infectious disease in a population of moving individuals. *Journal of Physics A: Mathematical and General* 26:3707-3717.
- Clark, A. B., and A. B. Lawson. 2002. Spatio-temporal clustering of small area health data. Pages 235-256. *In* A. B. Lawson, and D. Denison, Editors. *Spatial cluster modeling*. Chapman & Hall, London, England.
- Conner, M. M., and M. W. Miller. 2004. Movement patterns and spatial epidemiology of a prion disease in mule deer population units. *Ecological Applications* 14:1870-1881.
- Cressie, N. A. C. 1993. *Statistics for spatial data*. John Wiley & Sons, Inc., New York, New York, USA.
- Cross, P. C., J. O. Lloyd-Smith, and W. M. Getz. 2005. Disentangling association patterns in fission-fusion societies using African buffalo as an example. *Animal Behaviour* 9:499-506.
- Diggle, P. J. 2000. Overview of statistical methods for disease mapping and its relationship to cluster detection. Pages 87-103 *In* P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, Editors. *Spatial epidemiology*. Oxford University Press, Oxford, England.
- Diggle, P., P. Riberio, Jr., and O. Christensen. 2003. An introduction to model-based geostatistics. Pages 43-86 *In* spatial statistics and computational methods. J. Moller, Editor. Springer, New York, New York, USA.
- Elliott, P. J. Wakefield, N. Best, and D. Briggs, editors. 2001. *Spatial epidemiology methods and applications*. Oxford University Press, Oxford, England.
- Farnsworth, M. A., J. A. Hoeting, N. T. Hobbs, and M. W. Miller. 2006. Linking chronic wasting disease to mule deer movement scales: a hierarchical Bayesian approach. *Ecological Applications* 16:1026-1036.
- Hanski, I., and M. Gilpin. 1991. Metapopulation Dynamics: brief history and conceptual domain. Pages 3-16 *In* M. Gilpin and I. Hanski, Editors. *Metapopulation dynamics: empirical and theoretical investigations*. Academic Press, London, England.
- Heisey, D. M., D. O. Joly, and F. Messier. 2006. The fitting of general force-of-infection models to wildlife disease prevalence data. *Ecology* 87:2356-2365.
- Hess, G. R. 1996. Disease in metapopulation models: implications for conservation. *Ecology* 77:1617-1632.
- Hess, G. R., S. E. Randolph, P. Arneberg, C. Chemimi, C. Furlanello, J. Harwood, M. G. Roberts, and J. Swinton. 2002. Spatial aspects of disease dynamics. Pages 102-118 *In* P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heesterbeek, and A. P. Dobson, Editors. *The ecology of wildlife diseases*. Oxford University Press, Oxford, England.
- Jain, A. K., M. N. Murty, and P. J. Flynn. 1999. Data Clustering: A Review. *ACM Computing Surveys*, Vol. 31, No. 3.
- Joly, D. O., M. D. Samuel, J.A. Langenberg, J. A. Blanchong, C. A. Batha, R. E. Rolley, D. P. Keane, and C. A. Ribic. 2006. Spatial-epidemiology of chronic wasting disease in Wisconsin white-tailed deer. *Journal of Wildlife Diseases* 42:578-588.
- Kulldorff, M., and N. Nagarwalla. 1995. Spatial disease clusters: detection and inference. *Statistics in Medicine* 14:799-810.
- Kulldorff, M. 1997 A spatial scan statistic. *Communications in Statistics: Theory and Methods* 26:1481-1496.
- Kulldorff, M., and Information Management Services,

- Inc. 2006. SaTScan™ v7.0: Software for the spatial and space-time scan statistics. <http://www.satscan.org/>.
- Lawson, A. B. 2001. Statistical methods in spatial epidemiology. John Wiley & Sons, Chichester, England.
- Lawson, A. B., and F. L. R. Williams. 2001. An introductory guide to disease mapping. John Wiley & Sons, West Sussex, England.
- Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15:237-240.
- Moore, D. A. 1999. Spatial diffusion of raccoon rabies in Pennsylvania, USA. *Preventative Veterinary Medicine* 40:19-32.
- Noble, J. V. 1974. Geographic and temporal development of plagues. *Nature* 250:726-728.
- Norstrom, M., Pfeiffer, D. U., and Jarp, J. 2000. A space-time cluster investigation of an outbreak of acute respiratory disease in Norwegian cattle herds. *Preventive Veterinary Medicine* 47:107-119.
- Okubo, A. 1980. Diffusion and Ecological Problems: Mathematical Models. Springer-Verlag, Berlin, Germany.
- Perez A.M., M.P. Ward, P. Torres, and V. Ritacco. 2002. Use of spatial scan statistics and monitoring data to identify clustering of bovine tuberculosis in Argentina. *Preventive Veterinary Medicine* 56:63-74.
- Reluga, T., J. Medlock, and A. Galvani. 2006. A model of spatial epidemic spread when individuals move within overlapping home ranges. *Bulletin of Mathematical Biology* 68:401-416.
- Rhodes, C. J., and R. M. Anderson. 1997. For disease spread, epidemic thresholds and vaccination in a lattice model of disease spread. *Theoretical Population Biology* 52:101-118.
- Samuel, M. D., D. O. Joly, M. A. Wild, S. D. Wright, D. L. Otis, R. W. Werge, and M. W. Miller. 2003. Surveillance strategies for detecting chronic wasting disease in free-ranging deer and elk. USGS-National Wildlife Health Center Report. Madison, Wisconsin, USA. Available at: http://www.nwhc.usgs.gov/publications/fact_sheets/pdfs/cwd/CWD_Surveillance_Strategies.pdf.
- Sankoh O.A., Y. Ye, R. Sauerborn, O. Muller, and H. Becher. 2001. Clustering of childhood mortality in rural Burkina Faso. *International Journal of Epidemiology* 30:485-492.
- Shibli, S.A.R. 1997. Geostatistics FAQ. Available at: http://www.ai-geostats.org/fileadmin/Documents/FAQ/FAQ_Geostatistics_01.pdf
- Swinton, J. 1998. Extinction times and phase transitions for spatially structured closed epidemics. *Bulletin of Mathematical Biology* 60:215-230.
- Tango, T., and K. Takahashi. 2005. A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* 4:11.
- Viel, J. F., P. Arveux, J. Baverel, and J.Y. Cahn. 2000. Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *American Journal of Epidemiology* 152: 13-19.
- Wakefield, J. C., J. E. Kelsall, and S. E. Morris. 2000. Clustering, cluster detection, and spatial variation in risk. Pages 128-152 *In* P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, Editors. *Spatial epidemiology*. Oxford University Press, Oxford, England.